AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 112:5–16 (2000)

Georgian and Kurd mtDNA Sequence Analysis Shows a Lack of Correlation Between Languages and Female Genetic Lineages

DAVID COMAS,[1] FRANCESC CALAFELL,[1] NINA BENDUKIDZE,[2] LOURDES FAÑANÁS,[3] AND JAUME BERTRANPETIT[1]*
[1]*Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, 08003 Barcelona, Catalonia, Spain*
[2]*Institute for Clinical and Experimental Medicine, 14000 Prague, Czech Republic*
[3]*Unitat d'Antropologia, Departament de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Catalonia, Spain*

ABSTRACT Mitochondrial DNA sequences from Georgians and Kurds were analyzed in order to test the possible correlation between female lineages and languages in these two neighboring West Eurasian groups. Mitochondrial sequence pools in both populations are very similar despite their different linguistic and prehistoric backgrounds. Both populations present mtDNA lineages that clearly belong to the European gene pool, as shown by 1) similar nucleotide and sequence diversities; 2) a large number of sequences shared with the rest of European samples; 3) nonsignificant genetic distances; and 4) classification of the present lineages into the major European mtDNA haplogroups already described. The outlier position of the populations from the Caucasus according to classical genetic markers is not recognized in the present Georgian mtDNA sequence pool. This result suggests that the differentiation of mtDNA sequences in West Eurasia and the outlier features of Caucasian populations should be attributed to different processes. Moreover, the putative linguistic relationship between Caucasian groups and the Basques, another outlier population within Europe for classical genetic markers, is not detected by the analysis of mtDNA sequences. Am J Phys Anthropol 112:5–16, 2000.

The population history of humans has deeply affected the distribution of genetic variation, but has also determined to a large extent the linguistic differences among human populations. The parallel study of both, plus the insights contributed by archeology and paleoanthropology, can shed light on the past history of humans (Cavalli-Sforza et al., 1988, 1994). A case in point is Western Eurasia (i.e., Europe and the Middle East). It has been suggested that after the technologies for food production were discovered in the Neolithic, the populations of the Fertile Crescent expanded throughout Europe, largely replacing previously settled hunting-gathering groups (Ammermann and Cavalli-Sforza, 1984). This demographic process would have molded most of the autosomal genetic landscape in Europe (Sokal et al., 1991; Cavalli-Sforza et al., 1993; Piazza, 1993), but could also have brought Indo-European (IE) languages to Europe (Renfrew, 1987), and, in other directions, it would have caused the expansion of Elamo-Dravidian languages eastward to the Indian subcontinent and of Afro-Asiatic lan-

Grant sponsor: Dirección General de Investigación Científico-Técnica, Spain; Grant number: PB95-0267-C02-01; Grant sponsor: Direcció General de Recerca, Generalitat de Catalunya; Grant number: 1998SGR00009.

*Correspondence to: Jaume Bertranpetit, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Catalonia, Spain. E-mail: jaume.bertranpetit@cexs.upf.es

Received 11 June 1999; accepted 24 January 2000.

guages westward to North Africa (Renfrew, 1991; Barbujani and Pilastro, 1993).

We have studied two populations from West Eurasia, Georgians and Kurds, which are geographically proximate but which may have quite different population histories. Georgians are the main ethnic group in Georgia, south of the Caucasus range, and they speak a non-Indo-European language, often called Kartvelian, belonging to the isolated Kartvelian or South Caucasian family of languages. It has been suggested that the Neolithic wave of advance did not reach the Caucasus, and that food production either developed locally or was introduced through cultural contacts with farming groups, or both (Renfrew, 1991). Limited demographic changes, if any, associated with the origins of agriculture in the Caucasus could explain the persistence of a large linguistic diversity, including two local non-IE families of languages (North Caucasian and Kartvelian, nowadays considered as distinct families). The patterns of genetic variation in the Caucasus, including Georgia, as depicted by *classical* markers (i.e., blood groups and protein electromorphs), do not seem compatible with Neolithic demic diffusion (Barbujani et al., 1994). A similarity between Basque and Caucasian languages has been repeatedly suggested (Lafon, 1951; Gamkrelidze and Ivanov, 1990; Ruhlen, 1991). Both groups are linguistic isolates and it is possible that both languages are the remnants of those spoken in Europe before the expansion of the Indo-European languages (Renfrew, 1987; Barbujani and Pilastro, 1993). The putative relationship between languages and genes in Basques and the Caucasus might be attributed to remnant Upper Paleolithic features which might have not been deeply affected by the Neolithic wave of advance. A genetic test of this hypothesis, based on classical markers, did not show any particular genetic link between Caucasians and Basques (Bertorelle et al., 1995). Georgians seem to be genetically differentiated in relation to European populations, again as shown by classical markers (Nasidze and Salamatina, 1996).

Kurds inhabit primarily the highlands between the modern states of Turkey, Syria, Iraq, and Armenia, and migrated in more recent times to Iran and to Georgia. Kurds migrated to the Caucasus from Turkey and Iran at various times (mostly in the 19th century, but also during World War I) and have kept a distinct ethnic identity (Wixman, 1984).They speak an IE language from the Indo-Iranian group, closely related to Farsi (Persian). They may represent the descendants of the first shepherds that occupied the Kurdistan highlands since the first Neolithic. Previous genetic studies with classical markers (Cavalli-Sforza et al., 1994) have shown that Kurds are close to other Middle Eastern populations, but that they present higher genetic distances to Europeans.

Allele frequencies and DNA data, especially mtDNA control region sequences, have been shown to be poorly correlated. Examples are populations such as Basques (Bertranpetit and Cavalli-Sforza, 1991; Calafell and Bertranpetit, 1994), Finns (Cavalli-Sforza et al., 1994), Icelanders (Cavalli-Sforza et al., 1994), or Sardinians (Cavalli-Sforza et al., 1994; Cappello et al., 1996), whose mtDNA sequences resemble those found in most other European regions, despite their highly differentiated allele frequencies (Di Rienzo and Wilson, 1991; Bertranpetit et al., 1995; Sajantila et al., 1995; Richards et al., 1996). The causes of the apparent discrepancy of both types of information are the subject of a lively debate (Sajantila et al., 1995; Richards et al., 1996; Cavalli-Sforza and Minch, 1997; Barbujani et al., 1998).

We sequenced a 360-base-pair stretch in hypervariable segment I of the mtDNA control region. This allowed us to: 1) test the possible correlation between languages and female lineages using two West Eurasian populations with very different linguistic and prehistoric backgrounds; 2) place those sequence pools in a broader European and West Asian context; 3) compare patterns of genetic differentiation as shown by autosomal and mtDNA markers; and 4) test whether the linguistic affinities suggested for Georgians and Basques have a reflection in mtDNA sequences. The comprehension of this specific case study helps in clarifying the genetic structure and origin of West Eurasian populations.

MATERIALS AND METHODS

Blood samples from 45 Georgians and 29 Kurds were collected in rural areas around the city of Tbilisi, Georgia. All donors were healthy, unrelated individuals with autochthonous ancestry. Appropriate informed consent was obtained from all participants in this study.

DNA was extracted from blood samples using standard methods. The first hypervariable segment of the mtDNA control region was amplified using primers L15996 and H16401 (Vigilant et al., 1991). The amplification products were purified with the GenClean (BIO 101) kit and the sequence reaction was performed on each strand, using the same amplification primers, with the DNA Sequencing Kit (Perkin-Elmer, Oak Brook, IL), Dye Terminator Cycle Sequencing with AmpliTaqR DNA Polymerase. The product of the sequence reaction was run in an ABI PRISM 377 (Perkin-Elmer) automatic sequencer, and the sequences were aligned using the ESEE computer program (Cabot, 1988). Sequences from position 16,024 to 16,383 are shown in Figure 1 and are available upon request from david.comas@cexs.upf.es.

Statistical analysis

For classification purposes, a median network (Bandelt et al., 1995) of sequences was drawn. The groups of sequences found in the network were compared with the sequence motifs associated to high-resolution mtDNA RFLP haplogroups (Torroni et al., 1996, 1998; Macaulay et al., 1999).

mtDNA sequence data in Georgians and Kurds was compared to other European and West Asian populations: Basques (pooled from Bertranpetit et al., 1995; Côrte-Real et al., 1996), Spanish (pooled from Pinto et al., 1996; Côrte-Real et al., 1996), Tuscans (Francalacci et al., 1996), British (Piercy et al., 1993), Bavarians, Danes (Richards et al., 1996), Finns (Sajantila et al., 1995; Richards et al., 1996), Bulgarians (Calafell et al., 1996), Turks (pooled from Calafell et al., 1996; Comas et al., 1996; Richards et al., 1996), Adygei, Druze (Macaulay et al., 1999), and Middle Easterners (Di Rienzo and Wilson, 1991). For some analyses, an extended database was used, with European (Di Rienzo and Wilson, 1991; Handt et al., 1994; Pult et al., 1994; Sajantila et al., 1995; Richards et al., 1996; Stenico et al., 1996; Côrte-Real et al. 1996) and Asian sequences (Shields et al., 1993; Mountain et al., 1995; Horai et al., 1996; Kolman et al., 1996; Comas et al., 1998). Given the high number of undetermined positions, we selected 52 complete sequences of the 66 Taiwan Chinese sequences reported by Horai et al. (1996).

Intrapopulational mtDNA variation was mesured with two parameters. Sequence diversity (h) was calculated as $h = (n/n - 1)(1 - \Sigma x_i^2)$ where x_i is the sequence frequency and n is the number of individuals in the sample (Nei, 1987). Nucleotide diversity (π) was calculated as $\pi = (1/L) \times (n/n - 1)(\Sigma x_{ij}^2)$, where x_{ij} is the frequency of each of the four nucleotides at position j, L is the sequence length, and n is the sample size (Nei, 1987).

Genetic distances between populations were calculated using the mismatch-intermatch means of the pairwise difference distribution. Then $D = d_{ij} - (d_{ii} + d_{jj})/2$, where d_{ij} is the intermatch mean (i.e., the mean number of nucleotide differences) between populations i and j, and d_{ii} and d_{jj} are the mismatch values within populations i and j, respectively (Rao, 1982; Nei, 1987). In order to test whether the genetic distances found were significantly larger than 0, a randomization test was performed: the empirical null distribution of the genetic distance was obtained by bootstrapping 1,000 times over nucleotide positions and repeating the distance calculations. From this distribution, a 95% confidence interval was obtained for the distance, which would be significantly larger than zero if the confidence interval did not contain zero. The same bootstrap distance matrices were used to test whether the distance between two populations was significantly different from the distance between another pair of populations.

Neighbor-joining trees (Saitou and Nei, 1987) were built from the distance matrix, and the robustness of the nodes was estimated by the bootstrap method (Efron, 1982). The distance matrix was also repre-

```
                 1111111111111111111111111111111111111111111111111111111111111111
                 6666666666666666666666666666666666666666666666666666666666666666
                 0000001111111111111111112222222222222222222222223333333333333333
                 6678892244566677788889901222233444566777788999001111245555 6
                 7913636957336824623692398023424039601024847246491789432569 2

ANDERSON         CCCCTTTGGCGAACTCCAACTCCTCACCTCCATTCCCCAGCACCCCTATAAGTATCTTT  geo kur  hapl
GEO5             ...........................................................  7   6   H
GEO43            ..........................................T................  1   0
GEO49            ....................C......................................  1   0
KUR6             ..............................................C............  0   2
KUR4             ...........................................T.....C.....C....  0   1
KUR29            ........................T..................................  0   1
KUR22            ..............................................G..T........  0   1
KUR27            ......................T..............T......................  0   1
KUR11            ..T.....A......................T.............................  0   1

GEO85            .T....C.A.............................T......................  1   0   J
GEO119           .T....C.A.....C......T....T.........T......................  1   0

GEO102           ......C.....................................T..............  2   0   T
GEO34            ......C.................................A....T..............  1   0
GEO88            ......C....G........C...........C...........T..............  1   0
GEO72            ......C..A.G.......TC.........................T..............  2   0
GEO87            ......C.....................................TT.............  4   0
GEO79            ....C.C.....................................TT............C  1   0
KUR30            ......C............................T.........TT...G.........  0   1
KUR1             T.....C.....................................C...........  0   1

GEO40            ............................C...................C..........  2   1   K
GEO8             ......................C....C...................C..A.......  1   0
GEO24            .....C......................C...................C..........  2   0
KUR7             .....C......................C......T............C..........  0   2
KUR20            ............................C..G................C..........  0   1
KUR21            ............................C.........G.........C..........  0   1

GEO20            .......A.......T.CC.C........A...C............C............  1   0   U1
KUR23            ................CC.C............C.........................  0   1

GEO46            ...................................................G.....  1   0   U3
GEO11            ............T.......T................................G.....  1   0
KUR15            ............-.....................................G.....  0   1
KUR8             ......................T.........C...................G.....  0   1

GEO42            .....................................................C..  3   0   U4
GEO60            ..........................................G..............C..  1   0
GEO66            ....C.....................................................C..  1   0

GEO41            ......................T............T..T.....................  2   0   U5
GEO104           ..................................T..T.....................  1   0

GEO75            ..........................................T...............  1   0   W
GEO103           ............................C...................T..........  1   0
GEO32            ..........................T.................T..............  1   0
KUR10            ......................T.....T..............T........C......  0   1

KUR17            ....................TC...................T..................  0   1   X
GEO68            ......................C......T............T..................  2   0

KUR14            ........A....-..............T.............................  0   1   I
GEO13            ........A............C......T.....C..............C........C.  1   0
KUR5             ...T...A......C.....C......T............................T...  0   1

KUR3             ....C.....A................T.......................A.......  0   1   Other
GEO84            ..........................T.............................C  1   0
KUR13            ........A........G..........T.............................  0   1
```

Fig. 1.

sented by means of principal-coordinate analysis (Gower, 1966; Cavalli-Sforza et al., 1994, p.42).

RESULTS

The hypervariable segment I sequences of the mtDNA were obtained for 45 Georgian and 29 Kurdish individuals, which showed a total of 28 and 22 different sequences, respectively (Fig. 1). The sequences of both samples were defined by 41 segregating nucleotide positions, 23 of which are shared between them. Most segregating sites were transitions, except for five transversions (sites 16147, 16220, and 16232 in Georgians; and sites 16176 and 16318 in Kurds) and for the deletion in two Kurds of one A in a run of five As from position 16162 to 16166. The most frequent nucleotide in all the segregating sites in both populations is always the one found in the Cambridge reference sequence (CRS) (Anderson et al., 1981), although some positions present a high degree of polymorphism: 28.9% of the Georgians present a C at position 16126; 24.4% of the Georgians present a T at position 16294; and, 31.3% of the Kurds present a C at position 16311. The CRS represents 15.6% of the Georgian and 20.7% of the Kurdish sequences.

Diversity parameters

Nucleotide diversity was estimated at 0.0127 in Georgians and at 0.0123 in Kurds; both values are similar to those found in European populations and lower than values found in the Near East (0.0145 in Druze, 0.0155 in Turks, and 0.0197 in Middle Easterners) and Central Asia (0.0164–0.0185; Comas et al., 1998; Kolman et al., 1996).

Differences in sequence diversity found in Georgians (h = 0.964) and Kurds (h = 0.958) were not statistically significant (P = 0.4059; tested by a nonparametric procedure described in Mateu et al., 1997), and those values were similar to those found in European populations but lower than those found in Turks (h = 0.988; P = 0.0532 against Georgians and P = 0.0442 against Kurds) and Middle Easterners (h = 0.995; P = 0.0004 against Georgians and P = 0.0012 against Kurds).

Sequence sharing

Both samples share only two sequences: the CRS, which was present in seven Georgians and in six Kurds; and a sequence with the substitutions 16224 C and 16311 C, found in two Georgians and one Kurd. A large number of sequences in both groups was described previously in European populations (15 out of 28 Georgian sequences and 4 out of 22 Kurdish sequences). When we focus on the Near East, Georgians share three sequences with Middle Easterners (one of which, GEO85, had been previously described only in Middle Easterners), four sequences with Druze, six sequences with Adygei, and nine sequences with Turks (one of which, GEO88, had been previously found only in Turks), whereas Kurds share two sequences with Middle Easterners, two sequences with Adygei, three sequences with Druze, and two with Turks. Sequence KUR22 was found in the Indian Havik and in the Central Asian Uighur, but neither in Europe nor in Eastern Asia; nonetheless it clearly belongs to the haplogroup H, which is the most frequent in Europe. The majority of sequences reported in the present study (52% of the different sequences described in both groups) have not been reported previously. Some of the sequences in Georgians and Kurds that did not have a match in the European database could have an Asian or African origin. One of these sequences (KUR3) bears the 16223T–16319A motif associated with Asian haplogroup A (Torroni et al., 1993), and sequence

Fig. 1. Polymorphic sites of the 48 different sequences found in the Caucasus. Sequences and base positions are given in comparison to the Cambridge reference sequence (CRS) (Anderson et al., 1981). Sequences are sorted by groups found in a median network (Bandelt et al., 1995). The geo and kur columns indicate the number of sequences found in Georgians and Kurds, respectively. The hapl column indicates tentative assignment to high-resolution RFLP mtDNA haplogroups, according to the control-region sequences reported by Macaulay et al. (1999) and Kivisild (personal communication). The group labeled "Other" contains sequences sharing 16223T and belonging probably to haplogroups A (KUR3), D (GEO84), and L3a* (KUR13).

GEO 84 presents the motif 16223T–16362C, associated with Asian haplogroup D (Torroni et al., 1993). Sequence KUR13 presents 16145A–16176G–16233T and 16390A (position not shown in Fig. 1), which differs in only one position to one Druze sequence classified within the L3a* haplogroup by RFLPs and HVSI (Macaulay et al., 1999). Aside from these few exceptions, we can conclude that almost all Georgian and Kurdish sequences belong to the European/West Asian mtDNA sequence pool.

We constructed a median network (Bandelt et al., 1995) with all Georgian and Kurdish sequences (not shown). It defined 12 different groups of sequences (Fig. 1), 11 of which matched the motifs associated with European mtDNA RFLP haplogroups (Torroni et al., 1996, 1998; Macaulay et al., 1999; T. Kivisild, personal communication). The haplogroup labeled H in Figure 1 contains the sequences closely related to the CRS that do not belong to any of the other haplogroups. Since position 73 of HVSII was not tested, some sequences included in this group could belong to haplogroup H or to a precursor of this haplogroup. Sequences included in haplogroup I presented an A at position 16391 (not shown in Fig. 1), which provides an RFLP-site loss for *Ava*II and RFLP-site gain for *Bam*HI and *Mbo*I according to Macaulay et al. (1999) and confirms its inclusion in haplogroup I. The group labeled "Other" in Figure 1 contains three sequences sharing 16223T that may belong to haplogroups A and D (both Asian) and L3a* (African). In total, 3 individuals out of 74 could bear non-West Eurasian motifs, but all other individuals constitute a typically European mtDNA sequence pool.

Pairwise difference distributions

Both Georgians and Kurds present bell-shaped nucleotide pairwise difference distributions. The mean nucleotide pairwise difference value is similar in both groups: 4.57 in Georgians, and 4.42 in Kurds, lower values than those found in the Near East (5.22 in Druze, 5.45 in Turks, and 7.08 in Middle Easterners), but closer to those found in most European populations, which range from 2.95 (Basques) to 5.03 (Tuscans; median, 4.22). Both populations showed, as do all European populations, an excess of segregating sites over mean pairwise differences. This is measured by the D of Tajima (1989), which was -1.773 in Georgians ($P < 0.10$) and -2.140 in Kurds ($P \approx 0.01$).

Georgians and Kurds in a European/West Asian context

Genetic distances were computed among Georgians, Kurds, and a selection of European (Basque, Spanish, Tuscan, Bavarian, British, Danish, Bulgarian, Adygei, and Finnish) and West Asian (Middle Eastern, Druze, and Turkish; see reference citations in Materials and Methods) populations. Georgians and Kurds showed short, and even negative genetic distances to Europeans and West Asians; we determined by a bootstrap method that the distances from Georgians to all Europeans and West Asians (except for Basques, Druze, and Middle Easterners) were not significantly different from 0 ($P > 0.05$), and that was also the case between Kurds and Bulgarians, Tuscans, and Georgians. The distance matrix was depicted as a neighbor-joining tree (Fig. 2), and the robustness of its branches was assessed through 1,000 bootstrap iterations. In the neighbor-joining tree, Georgians and Kurds are found amidst all European populations, in a section of the tree where bootstrap values are extremely low (2–41%), and there does not seem to be a population structure amenable to interpretation. The European cluster (including Georgians and Kurds) joins Turks, Druze, and Middle Easterners with the only reasonably robust branches (bootstrap values: 72% between Europeans and all Near Easterners, and 87% between Middle Easterners and Druze) in the tree.

A principal-coordinate (PC) analysis based on the distance matrix was also performed. Figure 3 shows the two-dimensional plot of the first two PC axes, which account for 69.3% of the variance observed. The first PC separates the Eastern populations (Middle Easterners and Druze) from the rest of populations, including the Georgians and the Kurds, placing the Basques on the opposite edge. The second PC separates Middle Easterners and Druze, whereas the rest of populations present intermediate values.

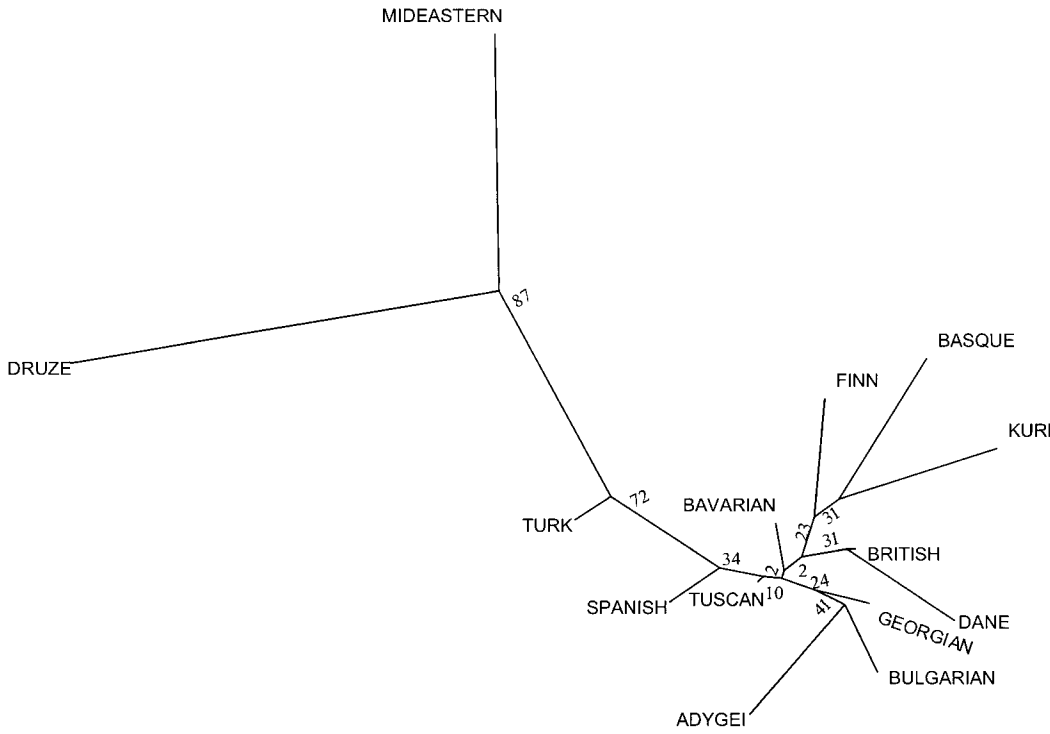

Fig. 2. Neighbor-joining tree among West Eurasian populations. Figures in cursive are percent bootstrap values, estimated from 1,000 iterations.

Relationship between Basques and Caucasians

Patterns of within-population sequence variation are different among Basques and populations from the Caucasus (Georgians and Adygei). Sequence diversity (0.936) and nucleotide diversity (0.0082) in Basques are reduced compared to the rest of the European populations, and this fact is also apparent in comparing them to the Georgians and the Adygei. Basques present also a lower mean pairwise difference value (2.95) compared to Georgians (4.57) and Adygei (5.08).

From the 52 different sequences described in 106 Basques (pooled from Bertranpetit et al., 1995; Côrte-Real et al., 1996), only five (i.e., GEO5, GEO49, GEO40, GEO87, and GEO68; Fig. 1) are shared with the present 28 different sequences found in Georgians, and four with the 31 different sequences found in Adygei, and all of them have been reported in many other European populations. Moreover, there are some important differences in the lineages present in those groups (Table 1):

1. Haplogroup H (and closely related sequences) is significantly more frequent in Basques than in the Caucasus (67% vs. 20% in Georgians, $P < 0.001$; and 67% vs. 40% in Adygei, $P = 0.002$, Fisher's exact text). It should be noted that the frequency in Basques is among the highest in Europe, while the frequency in Georgians is one of the lowest in Europe and the Middle East.
2. Haplogroup V, defined by a C in position 16298 (Torroni et al., 1998), was not found in the Caucasus.
3. Haplogroup T is more frequent in Georgians (24% vs. 6%; $P = 0.002$, Fisher's exact text) and in Adygei (14% vs. 6%; not statistically significant) than in Basques, and, in general, than in other European populations (0–13%).
4. U5 lineages, which bear 16270T, are always associated with 16256T in Geor-

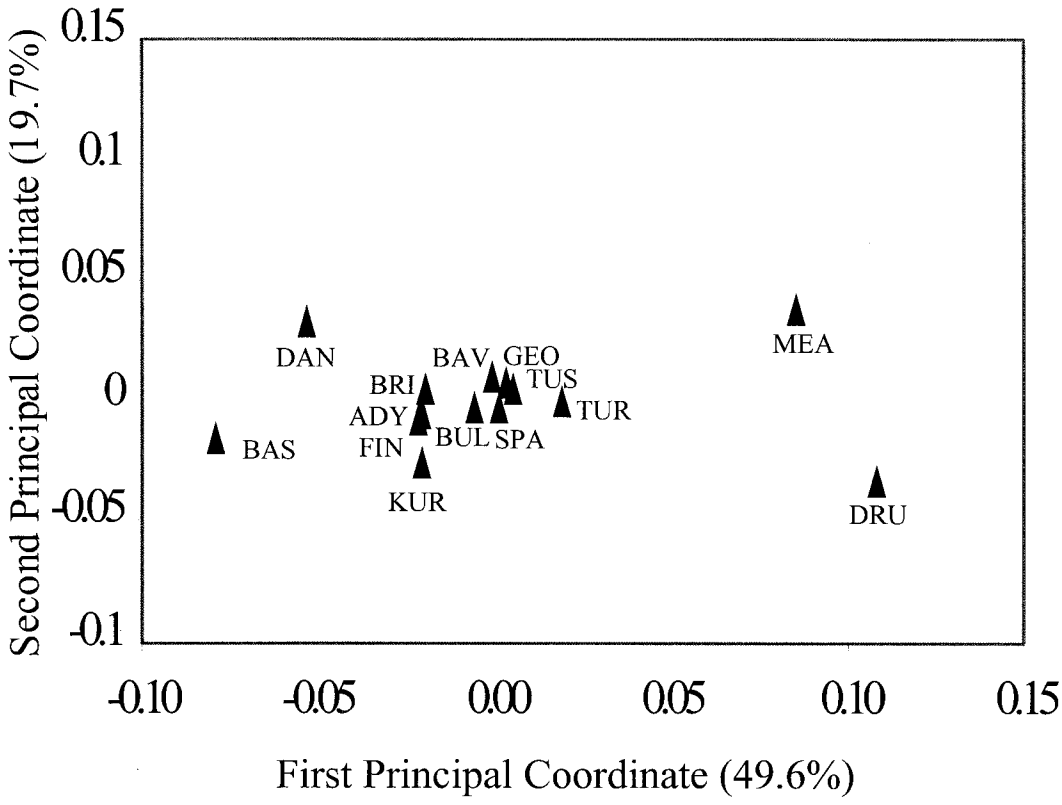

Fig. 3. Principal-coordinate (PC) analysis of the distance matrix among West Eurasian populations. The first PC explains 49.6% of the variation, and the second explains 19.7%. ADY, Adygei; BAS, Basque; BAV, Bavarian; BRI, British; BUL, Bulgarian; DAN, Danish; DRU, Druze; FIN, Finnish; GEO, Georgian; KUR, Kurdish; MEA, Middle Easterner; SPA, Spanish; TUR, Turkish; TUS, Tuscan.

gians and Adygei, which is not the case for the Basques.

5. Haplogroups U1, U2, U3, U4, I, W, C, and D were found in the Caucasus but not in Basques. For W ($P = 0.025$) and U4 ($P = 0.002$), the differences in frequency between Basques and Georgians were statistically significant; as were the differences in frequency between Basques and Adygei for U1 ($P = 0.032$), U3 ($P < 0.001$), and C ($P = 0.032$).

Overall, the genetic distance between Georgians and Basques (d = 0.0991) was significantly higher than 0 ($P = 0.0034$). Moreover, if any close link between Georgians and Basques existed, the genetic distance between both populations should be lower than the genetic distances between Georgians and other European populations. Taking the British (Piercy et al., 1993) as a

TABLE 1. ABSOLUTE HAPLOGROUP FREQUENCIES IN BASQUES, GEORGIANS, AND ADYGEI[1]

	Basque (N = 106)	Georgian (N = 45)	Adygei (N = 50)
H[2]	72 (67.9)	9 (20.0)	20 (40.0)
V	7 (6.6)	0	0
J	3 (2.8)	2 (4.4)	2 (4.0)
T	6 (5.7)	11 (24.4)	7 (14.0)
K	5 (4.7)	5 (11.1)	1 (2.0)
U1	0	1 (2.2)	3 (6.0)
U2	0	0	1 (2.0)
U3	0	2 (4.4)	7 (14.0)
U4	0	5 (11.1)	1 (2.0)
U5	11 (10.4)	3 (6.7)	4 (8.0)
I	0	1 (2.2)	0
W	0	3 (6.7)	1 (2.0)
X	2 (1.9)	2 (4.4)	0
C	0	0	3 (6.0)
D	0	1 (2.2)	0

[1] Haplogroups were inferred from control region sequences in Basques and Georgians, following Macaulay et al. (1999). In parentheses, percent frequencies.
[2] Since position 73 was not tested in Basques and Georgians, the present haplogroup H contains CRS and closely related sequences, which include haplogroup H and its precursors.

reference European population, the genetic distance between Basques and Georgians (0.0991) is significantly higher ($P = 0.016$) than the genetic distance between British and Georgians (0.0203). We can conclude that Georgians do not seem to present a closer relation to Basques than to the rest of European populations in the female lineage, as revealed by mtDNA sequences.

DISCUSSION

The Caucasus area shows allele frequencies for classical markers that are often outside the range observed for European populations (Barbujani et al., 1994; Nasidze and Salamatina, 1996). Nevertheless, the present analysis demonstrates that mtDNA lineages in Georgians and Kurds clearly belong to the European gene pool, with minor differences between them and with other European populations. This conclusion is supported by 1) the similar nucleotide and sequence diversities; 2) the large number of shared sequences; 3) the insignificant genetic distances; and 4) the presence of substitutions that allow one to attribute almost all sequences to European haplogroups (Macaulay et al., 1999). On the basis of mtDNA variation, the Kurdish and Georgian samples appear more closely related to European than to Middle Eastern samples; nucleotide and sequence diversities are higher in the Middle East and in Turkey than among Georgians and Kurds.

In spite of their geographical proximity, the population history of Georgians and Kurds may have been quite dissimilar. The Neolithic demic wave of advance may have not entered the Caucasus (Barbujani et al., 1994), where the food-producing technologies may have been local developments or may have been introduced by cultural diffusion (Renfrew, 1991). Conversely, the Kurds may be the descendants of the first shepherds that populated the Kurdistan highlands. In the view of Renfrew (1987), Indo-European (IE) languages would have diffused with the Neolithic wave of advance, and thus, the non-Indo-European Georgian language would be a pre-Neolithic relic. Our results show that mtDNA sequences do not reflect this historic pattern: Georgian and Kurdish mtDNA sequences are quite similar to each other and to European lineages. Moreover, there is little discernible structure in the European mtDNA gene pool: beyond linguistic (IE vs. non-IE-speaking populations) and geographic distances, the European mtDNA pool appears quite homogeneous. In this respect, Georgians and Kurds appear to conform to a general pattern: populations such as Finns, Basques, and Sardinians are genetic outliers for classical polymorphisms but are not distinct with respect to mtDNA sequences. The interpretation of this observation is the object of a lively debate (Sajantila et al., 1995; Richards et al., 1996; Cavalli-Sforza and Minch, 1997; Barbujani et al., 1998), concerning mainly the relative importance of the Upper Paleolithic vs. Neolithic events in the European genetic landscape. It should be noted, though, that mtDNA behaves as a single locus and that it is subject to the vagaries of genetic drift. Moreover, it has been suggested that variation in mtDNA may be under selective pressures (Excoffier, 1990; Nachman et al., 1996; Templeton, 1996; Wise et al., 1998). However, the amount of phylogenetic information contained in mtDNA sequences can still be used to infer population affinities as shown by a large number of works. A higher female-mediated migration rate has been invoked to explain the apparent homogeneity of matrilineal mtDNA sequences compared to nuclear markers in Europe (Cavalli-Sforza and Minch, 1997), and it is often found that Y-chromosome markers show higher interpopulation diversities than mtDNA sequences (Salem et al., 1996; Scozzari et al., 1997; Pérez-Lezaun et al., 1999), even across long-established linguistic barriers. Moreover, Poloni et al. (1997) showed that Y-chromosome variation correlates better with geography and linguistic affiliation than do female lineages. The results of this study do not agree with the view that the spread of what we call "European" mitochondrial alleles paralleled the introduction of IE languages in Eurasia. It is still unknown whether biparentally transmitted alleles, and alleles of the Y chromosome, may be distributed in a way that more closely matches the distribution of languages, as

observed by Poloni et al. (1997) in Africa and Europe.

In this context, it is not surprising that we failed to identify the remnants of a putative Paleolithic background shared between Basques and Georgians, as has been postulated from a linguistic standpoint (Lafon, 1951). However, we failed to identify any genetic connection between Basques and Georgians, in spite of a putative Neolithic migration postulated by Calderón et al. (1998) based on immunoglobulin typing in Basques (but not in Caucasians). In the present study, Basques and Georgians are at the extremes of the distribution of several mtDNA haplogroups in Europe/West Asia (high H and V frequencies in Basques, high T frequency in Georgians), and it is remarkable that the only statistically significant genetic distance between Georgians and a European population was found with Basques. Divergence by mutation and drift, and the fact that Basques and Georgians are separated by a large geographical distance, may have obscured any common mtDNA background, as was postulated for nuclear markers (Bertorelle et al., 1996).

As Darwin noted in Chapter 14 of *The Origin of the Species*, there is much to be learnt from the evolution of genes and languages. However, there may be even more information on population history in instances where genes and languages seem to tell different stories.

ACKNOWLEDGMENTS

We thank Dr. E. Meschishvili and Dr. R. Guladze for invaluable help in collecting the samples, and Dr. E. Ivaskova for encouragement. Thanks go to T. Kivisild, University of Tartu, for sharing unpublished data and assistance in the correspondence of sequences and haplogroups.

LITERATURE CITED

Ammerman AJ, Cavalli-Sforza LL. 1984. The Neolithic transition and the genetics of populations in Europe. Princeton, NJ: Princeton University Press.

Anderson S, Bankier AT, Barrell BG, De Bruijn MH, Coulson AR, Sanger F, Schreier PH, Smith AJH, Staden R, Young G. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Bandelt HJ, Forster P, Sykes B, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

Barbujani G, Pilastro A. 1993. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic family. Proc Natl Acad Sci USA 90:4670–4673.

Barbujani G, Whitehead GN, Bertorelle G, Nasidze IS. 1994. Testing hypotheses on processes of genetic and linguistic change in the Caucasus. Hum Biol 66:843–864.

Barbujani G, Bertorelle G, Chikhi L. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:488–491.

Bertorelle G, Bertranpetit J, Calafell F, Nasidze I, Barbujani G. 1995. Do Basque- and Caucasian-speaking populations share non-Indoeuropean ancestors? Eur J Hum Genet 3:256–263.

Bertranpetit J, Cavalli-Sforza LL. 1991. A genetic reconstruction of the history of the population of the Iberian Peninsula. Ann Hum Genet 55:51–67.

Bertranpetit J, Sala J, Calafell F, Underhill P, Moral P, Comas D. 1995. Human mitochondrial DNA variation and the origin of the Basques. Ann Hum Genet 59: 63–81.

Cabot EL. 1988. ESEE: the eyeball sequence editor, version 1.06. Burnaby: University of British Columbia.

Calafell F, Bertranpetit J. 1994. Principal component analysis of gene frequencies and the origin of Basques. Am J Phys Anthropol 93:201–215.

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35–49.

Calderón R, Vidales C, Pena JA, Miranda A, Dugoujon JM. 1998. Immunoglobulin allotypes (GM and KM) in Basques from Spain: approach to the origin of the Basque population. Hum Biol 70:667–698.

Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, Piazza A. 1996. Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125–141.

Cavalli-Sforza LL, Minch E. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 61:247–251.

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain JL. 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. Proc Natl Acad Sci USA 85:6002–6006.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. Science 259:639–646.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. History and geography of human genes. Princeton, NJ: Princeton University Press.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bertranpetit J. 1996. Geographic variation in human mtDNA control region sequence: the population history of Turkey and its relationship to European populations. Mol Biol Evol 13:1067–1077.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimón J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J. 1998. Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. Am J Hum Genet 63:1824–1838.

Côrte-Real HBSM, Macaulay V, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha SS, Bertranpetit J, Sykes B. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350.

Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601.
Efron B. 1982. The jacknife, the bootstrap and other resampling plans. Philadelphia: Society for Industrial and Applied Mathematics.
Excoffier L. 1990. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J Mol Evol 30: 125–139.
Francalacci P, Bertranpetit J, Calafell F, Underhill P. 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100: 443–460.
Gamkrelidze TV, Ivanov VV. 1990. The early history of Indo-European languages. Sci Am 262:82–88.
Gower JC. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.
Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, Schaffner W, Utermann G, Sykes B, Pääbo S. 1994. Molecular genetic analysis of the Tyrolean Ice Man. Science 264:1775–1778.
Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH. 1996. MtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590.
Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334.
Lafon R. 1951. Concordances morphologiques entre le basque et les langues caucasiques. Word 7:227–244.
Macaulay V, Richards M, Hikey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synethesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249.
Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Bertranpetit J. 1997. A tale of two islands: population history and and mitochondrial DNA sequences of Bioko and Sâo Tomé, Gulf of Guinea. Ann Hum Genet 61:507–518.
Mountain JL, Hebert SS, Bhattacharyya P, Underhill P, Ottolenghi C, Gadgil M, Cavalli-Sforza LL. 1995. Demographic history of India and mitochondrial DNA sequence diversity. Am J Hum Genet 56:979–992.
Nachman MW, Brown WM, Stoneking M, Aquadro CF. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142:953–963.
Nasidze IS, Salamatina NV. 1996. Genetic characteristics of the Georgian population. Gene Geogr 10:105–112.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
Pérez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martínez-Arias R, Clarimón J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J. 1999. Sex-specific migration patterns in Central Asian populations revealed by the analysis of Y-chromosome STRs and mtDNA. Am J Hum Genet 65:208–219.
Piazza A. 1993. Who are the Europeans? Science 260: 1767–1769.
Piercy R, Sullivan KM, Benson N, Gill P. 1993. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Leg Med 106:85–90.
Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60:321–330.
Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L. 1997. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015–1035.
Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S. 1994. Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837–840.
Rao CR. 1982. Diversity and dissimilarity coefficients: a unified approach. Theor Pop Biol 21:24–43.
Renfrew C. 1987. Archaeology and language. The puzzle of Indo-European origins. London: Jonathan Cape.
Renfrew C. 1991. Before Babel: speculations on the origins of linguistic diversity. Camb Archaeol J 1:3–23.
Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203.
Ruhlen M. 1991. A guide to the world's languages, 2nd ed. Stanford: Stanford University Press.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.
Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, Di Rienzo A, Pääbo S. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. Genet Res 5:45–52.
Salem AH, Badr FM, Gaballah MF, Pääbo S. 1996. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. Am J Hum Genet 59:741–743.
Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, Wallace DC, Kidd KK, Olckers A, Moral P, Terrenato L, Akar N, Qamar R, Mansoor A, Mehdi SQ, Meloni G, Vona G, Cole DEC, Cai W, Novelletto A. 1997. Differential structuring of human populations for homologous X and Y microsatellite loci. Am J Hum Genet 61:719–733.
Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Redd JK, Ward RH. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562.
Sokal RR, Oden NL, Wilson L. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 351:143–145.
Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio MA, Corrain C, Barbujani G. 1996. High mitochondrial sequence diversity in two linguistic isolates of the Alps. Am J Hum Genet 59:1363–1375.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.
Templeton AR. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics 144:1263–1270.
Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850.

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonné-Tamir B, Scozzari R. 1998. mtDNA analysis reveal a major late Paleolithic population expansion from Southwestern to Northeastern Europe. Am J Hum Genet 62:1137–1152.

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1991. Mitocohondrial DNA sequences in single hairs from a Southern African population. Proc Natl Acad Sci USA 86:9350–9354.

Wise CA, Sraml M, Easteal S. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics 148:409–421.

Wixman R. 1984. The peoples of the USSR. An ethnographic handbook. Armonk, NY: M.E. Sharpe, Inc.